

一种基于 Hopfield 网络的立体匹配方法

胡海峰 熊银根

(中山大学电子与通信工程系, 广州 510275)

摘要 立体匹配是计算机视觉研究的经典难题,其算法的复杂度和精度直接影响了视觉系统对外部景物的重建性能。为此提出了一种新的基于神经网络的立体匹配方法,其基本思想是:在实现核线重排的前提下,利用唯一性、相容性以及相似性等匹配约束条件,建立反映对应极线间所有匹配点约束关系的能量函数,将其映射到二维 Hopfield 网络进行极小化求解,网络最后的稳态表示匹配点的对应关系;通过对图中所有极线进行上述操作,可以得到所求的视差图。与传统方法相比,本算法具有两个明显的特点:(1) 匹配单元采用了普通的图像点,可以直接获得稠密的深度图;(2) Hopfield 网的外部输入不再为常数,而是一个反映对应点灰度相似性关系的值。通过对合成图以及真实图景进行测试,验证了该方法的有效性。

关键词 立体匹配 Hopfield 网络 稠密深度图 灰度相似性

中图分类号: TP183 文献标识码: A 文章编号: 1006-8961(2004)06-0729-08

A New Stereo Matching Approach Based on Hopfield Network

HU Hai-feng, Xiong Yin-gen

(Electronics Engineering Dept., Zhongshan Univ., Guangzhou 510275)

Abstract Stereo matching is one of the classic difficult problems in the computer vision, and its complexity and precision hedge the capability of vision system to reconstruct the 3-D scene. This paper presents a new matching method based on neural network. On the condition that stereo rectification has been performed, the energy function is built on the basis of uniqueness, compatibility and similarity constraints, which reflects the constraint relations of all matching units of the same lines. It is then mapped onto a 2-D neural network for minimization, whose final stable state indicates the possible correspondence of the matching units. The depth map can be acquired through performing the above operation on the all epipolar lines. The algorithm has two traits relative to the traditional approach. 1. Individual pixel point but not scene point or edge line is adopted as matching unit and dense depth map could be obtained directly. 2. The external input of the nodes is not constant again and is the function of gray similarity of correspondent points. The experiments on the synthetic and real images demonstrate the feasibility of our approach.

Keywords stereo matching, Hopfield network, dense depth map, gray similarity

1 引言

立体匹配是计算机视觉研究的一个难点,其目的在于找出不同成像面对应点的关系并得到视差值,进而由求得的深度信息构建出景物的三维外观。匹配问题的解决对于机器人、遥感应用及物体识别等领域的研究具有重要的意义。

立体匹配是一个极难解决的问题,原因是在二

维空间搜索所有的对应点,计算量是惊人的,而且由于诸如噪声、光照变化、遮挡以及透视畸变等因素的影响,匹配精度难以得到保证。为了降低运算量和提高精度,一个常用的方法是采用特征匹配^[1],即预先对原始图像进行特征提取,再对提取的特征属性进行匹配,最后利用匹配结果内插进行景物重建。这种方法对于对比度和明显光照变化等相对稳定,同时由于它只是通过对特征属性的简单比较而实现,因此速度也较快,但也存在以下不足:(1) 特征匹配通

常采用相关法、松弛法或其他方法得到,使用的是启发式知识,如图像平面上的刚性变换,因此不是鲁棒的;(2)从图像所提取的点常常不够精确。

对应点问题可以看作是代价函数的优化问题。在约束条件下,建立以基元匹配的相似性条件及相容性条件为基础的最优准则,然后进行点集的最优匹配搜索,使各基元的相似性与相容性达到总体最优。常用的最优搜索方法有动态规划方法、最大流法、Hopfield 网络(HNN)优化法以及遗传算法等,近年来进行匹配求解的优化方法可以参见文献[2]。

利用 Hopfield 网络^[3]进行立体匹配由来已久,早期的如文献[4]、[5]等,至今仍有一些相关论文的发表,如文献[6]等。这些方法基本思路相似,即先对图像进行特征点提取,然后利用特征点之间的位置信息以及唯一性约束来构造代价函数,并将其转化为 Lyapunov 函数的形式,用 Hopfield 网络进行极小化求解,神经元的输出结果表示特征点之间的对应关系。这些方法的好处在于匹配的速度快且精度高,能量函数可以显式地反映出匹配所固有的约束。不足之处在于:未考虑匹配基元的灰度相似关系;匹配所得为稀疏的视差图,在实用过程中还需进行内插,增加了重建的复杂度。

本文对传统的方法作了改进,体现在以下 3 个方面:(1)匹配基元采用的是普通像素点,而不是传统的特征点,因此可以直接得到稠密的深度图;(2)在构造能量函数的时候,引入了灰度相似性假设,因此节点的外部输入不再是常数,而为反映光照相似性的函数;(3)引入视差梯度作为描述对应点相容关系的参量。

2 基本原理

2.1 立体视觉的基本原理

图 1 所示为一个立体视觉系统, C_1, C_2 为摄像机, P 为空间点, p_1, p_2 为点 P 在两个图像平面的投影点,它们也互称为对应点。从图中可以看出点 P 为 $O_1 p_1$ 和 $O_2 p_2$ 的交点,因此如果能找出点 p_1, p_2 的对应关系(即实现对应点匹配),再利用系统已知的定标参数,就可以唯一确定 P 的空间位置了,这就是立体视觉的基本原理。

2.2 对应点问题转化为优化问题

优化问题可视为一个二元组 (F, C) ,其中 F 是

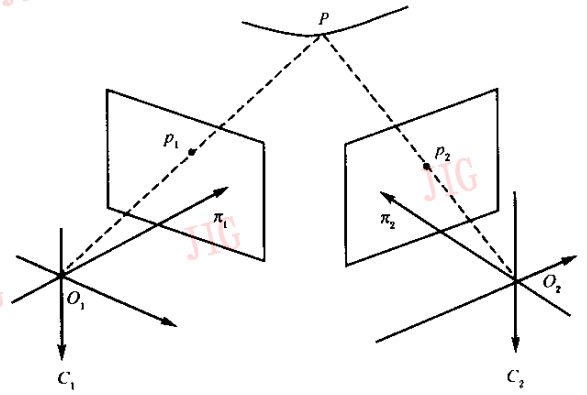


图 1 双目视觉系统

一个集合或可行点的定义域, C 是代价函数(目标函数),则映射 $C: F \rightarrow R$ 问题是寻找一个 $f \in F$,使得对一切 $y \in F$,有

$$C(f) \leq C(y) \quad (1)$$

这样一个 f 称为给定例子的优化解。

立体匹配问题可视为一个组合优化问题,其中 F 表示为某一对应关系全体

$$\{((u_1^l, v_1^l), (u_1^r, v_1^r)), ((u_2^l, v_2^l), (u_2^r, v_2^r)), \dots, ((u_N^l, v_N^l), (u_N^r, v_N^r))\}$$

其中, (u, v) 表示图像点的坐标, N 为对应点的数目,上标 l, r 分别表示左右图像,则对应问题的整体目标函数可以写为^[4]

$$J = \sum_{i=1}^k \lambda_i \left[\sum_{j=1}^N C_i((u_i^l, v_i^l), (u_i^r, v_i^r)) \right] \quad (2)$$

其中, C 表示为由匹配约束构造的代价函数, k 为约束数, N 为匹配的点数, λ_i 为足够大的常数,取值可以互不相同。

2.3 利用连续 HNN 求解优化问题

HNN 是一种常见的优化方法。它有 2 种基本模型,一种是离散的随机模型(DHNN),另一种是连续的确定模型(CHNN),其动力学系统方程为^[3]

$$\begin{cases} c_i \frac{du_i(t)}{dt} = \sum_{j=1}^n w_{ij} V_j(t) - \frac{u_i(t)}{R_i} + I_i \\ V_i(t) = g_i[u_i(t)], i = 1, 2, \dots, n \end{cases} \quad (3)$$

其中, w_{ij} 为连接权值, $u_i(t)$ 和 $V_i(t)$ 分别表示在 t 时刻,第 i 个神经元的输入和输出值; c_i 和 I_i 为可调节参数; $g(\cdot)$ 为神经元的非线性作用函数,通常为 sigmoid 函数:

$$V_i = g(u_i) = \frac{1}{2} \left[1 + \tanh \frac{u_i}{u_0} \right] \quad (4)$$

u_0 控制函数的斜率,当 $u_0 \rightarrow 0$ 时, g_i 变为阶跃函数。

CHNN 用于优化求解的基本思想是:将系统的能量函数与问题的目标函数对应,随机给系统一个初始状态 $s(0)$,让系统演化到稳定状态就可得到解。在实际中,如何把问题的目标函数表达成下述二次型的能量函数是一个关键^[3]:

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} V_i V_j - \sum_{i=1}^n I_i V_i$$

$$= -\frac{1}{2} \mathbf{x}^T \mathbf{W} \mathbf{x} - \mathbf{x}^T \mathbf{I} \quad (5)$$

其中, \mathbf{W} 为对角线全为 0 的对称矩阵, \mathbf{I} 为外加输入向量。

3 基于 CHNN 的立体匹配算法

图 2 所示为采用的 Hopfield 网络的拓扑结构,它可以视为一个 $N_l \times N_r$ 的二维矩阵, N_l 和 N_r 分别表示在左右对应极线上进行匹配的像点个数。由于采用普通的像素点作为匹配基元,因此有 $N_l = N_r = N_w$, N_w 表示图像的宽度。神经元 $n(x, i)$ (简称为 n_{xi}) 表示左极线上序号为 x 的像点与右极线上序号为 i 的像点之间的对应关系。 I_{xi} 和 I_{yj} 分别表示 n_{xi} 和 n_{yj} 的外部输入, W_{xij} 为它们之间的连接强度, V_{xi} 和 V_{yj} 表示为 n_{xi} 和 n_{yj} 的输出。

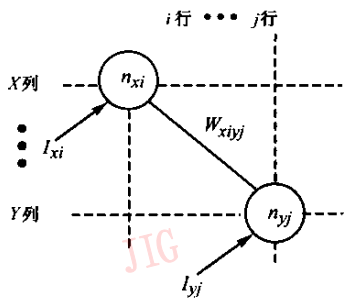


图 2 采用的二维 HNN 拓扑结构

3.1 构造代价函数

为了实现最优匹配,代价函数应该反映匹配固有的约束,即唯一性约束、平滑性约束以及相似性约束,可以认为 HNN 的动态变化过程是多个约束的极小化实现过程。

(1) 唯一性约束

对于一幅图中的任意点,在其他图中只有唯一的对应点。该约束可以写成

$$E_1 = \sum_{x=1}^{N_w} \sum_{i=1}^{N_w} \sum_{j=1, j \neq i}^{N_w} V_{xi} V_{xj} + \sum_{x=1}^{N_w} \sum_{i=1}^{N_w} \sum_{y=1, y \neq x}^{N_w} V_{xi} V_{yi} + \left(\sum_{x=1}^{N_w} \sum_{i=1}^{N_w} V_{xi} - N_w \right)^2 \quad (6)$$

式(6)表明,当一行(或列)中只存在一个激活神经元时, E_1 才趋于最小。

(2) 平滑性约束

视差大小反映了空间点的深度,由于实际场景中物体表面总是连续的,这就要求视差图应该是局部平滑的。视差梯度可以作为视差变化的一个测度,定义为:若左图像 π_1 中的点 p_1 与点 q_1 对应于右图像 π_2 中的点 p_2 与点 q_2 ,且点对 (p_1, q_1) 和 (p_2, q_2) 在二维矩阵中的坐标为 (x, i) 和 (y, j) ,则视差梯度可写为^[7]

$$G_d = 2 \frac{|(y-x) - (j-i)|}{|(y-x) + (j-i)|} \quad (7)$$

心理实验表明^[8],点 p_1 与 p_2 及 q_1 与 q_2 的匹配如果是相容的,则视差梯度应小于等于 2。

利用视差梯度,定义反映视差图平滑特性的相容性因子 C_{xij}

$$C_{xij} = 2 \exp \left(-\frac{(G_d - G_0)^2}{\lambda^2} \right) - 1 \quad (8)$$

其中, G_d 为视差梯度, λ, G_0 为控制参数。 λ 控制了函数变化快慢程度, G_0 则控制了函数经过最大值的位置,表示在视差梯度为 G_0 时相邻匹配点的相容性达到最佳。理想情况下 $G_0 = 0$,但考虑噪声和畸变的影响, G_0 一般为较小的常数。

相容性约束定义如下:

$$E_2 = - \sum_{x=1}^{N_w} \sum_{i=1}^{N_w} \sum_{y=1}^{N_w} \sum_{j=1}^{N_w} C_{xij} V_{xi} V_{yj} \quad (9)$$

(3) 相似性约束

由于论文直接将单个像素点作为匹配基元,此时能利用的只是像素的灰度信息。采用 Cox^[9]的观点,即对应点的灰度按正态分布概率趋于某一真值,用数学形式表达为:若 X 表示空间点, $p_k (k=1, 2)$ 为图像点, z_{p_k} 表示图像 k 中点的灰度信息,则 p_k 为 X 投影点的概率可定义如下^[9]:

$$P(p_k | X) = \left| (2\pi) S_k \right|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} (z - z_{p_k}) S_k^{-1} (z - z_{p_k}) \right\} \quad (10)$$

其中, S_k 为与偏差 $(z - z_{p_k})$ 有关的协方差, z 为某一未知的真值,可求得其近似值 \hat{z} 为^[9]

$$\hat{z} \approx z = \frac{S_{p_1}^{-1} z_{p_1} + S_{p_2}^{-1} z_{p_2}}{(S_{p_1}^{-1} + S_{p_2}^{-1})} \quad (11)$$

其中, S_{p_k} 是与 z_{p_k} 有关的协方差。

如果点对 (p_1, p_2) 在二维矩阵中的坐标为 (x, i) , 则由上述两式, 可定义相似性因子 S_{xi}

$$S_{xi} = P((x, i) | X) = P(x | X)P(i | X) \quad (12)$$

相似性约束可以写为

$$E_3 = - \sum_{x=1}^{N_w} \sum_{i=1}^{N_w} S_{xi} V_{xi} \quad (13)$$

3.2 能量函数

由上述约束函数, 可得到能量函数的形式为

$$\begin{aligned} E &= \frac{a}{2} E_1 + \frac{b}{2} E_2 + c E_3 \\ &= \frac{a}{2} \sum_{x=1}^{N_w} \sum_{i=1}^{N_w} \sum_{j=1, j \neq i}^{N_w} V_{xi} V_{xj} + \\ &\quad \frac{a}{2} \sum_{x=1}^{N_w} \sum_{i=1}^{N_w} \sum_{y=1, y \neq x}^{N_w} V_{xi} V_{yi} + \\ &\quad \frac{a}{2} \left(\sum_{x=1}^{N_w} \sum_{i=1}^{N_w} V_{xi} - N_w \right)^2 - \\ &\quad \frac{b}{2} \sum_{x=1}^{N_w} \sum_{i=1}^{N_w} \sum_{y=1}^{N_w} \sum_{j=1}^{N_w} C_{xijy} V_{xi} V_{yj} - \\ &\quad c \sum_{x=1}^{N_w} \sum_{i=1}^{N_w} S_{xi} V_{xi} \end{aligned} \quad (14)$$

其中, a, b, c 为正常数。

将式(14)与式(5)进行比较, 可得

$$W_{xijy} = -a(\delta_{xy} + \delta_{ij} - 2\delta_{xy}\delta_{ij}) + bC_{xijy}(1 - \delta_{xy})(1 - \delta_{ij}) \quad (15)$$

$$I_{xi} = aN_w + bS_{xi} \quad (16)$$

其中,

$$\delta_{xy} (\text{或 } \delta_{ij}) = \begin{cases} 1 & \text{如果 } x = y (\text{或 } i = j) \\ 0 & \text{其他} \end{cases} \quad (17)$$

3.3 能量函数的改进

实拍景物总是存在一定景深, 因此匹配时的视差值也是有范围的, 即用于实际匹配的神经元只是一小部分。图3显示了这些神经元所在的位置, $D = [d_{\min} \ d_{\max}]$ 为视差范围, 两条虚线将神经元分为2个部分, 虚线内部(黑实心点所在范围)为用于匹配的神经元, 记为

$$N_m: \{n_{xi} | 1 \leq x \leq N_w, x + d_{\min} \leq i \leq x + d_{\max}\} \quad (18)$$

若将矩阵坐标作一个简单平移 $T: (x, i) \rightarrow (x', i')$

$$\begin{cases} x' = x \\ i' = i - x - d_{\min} \end{cases} \quad (19)$$

则所有进行匹配的神经元在一个矩形范围内, 记为

$$N_m: \{n_{x'i'} | 1 \leq x' \leq N_w, 1 \leq i' \leq D_w\} \quad (20)$$

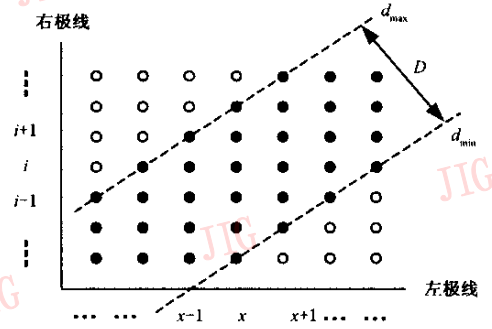


图3 实际进行动态更新的神经元(标为黑实心点)

其中, $D_w = d_{\max} - d_{\min} + 1$ 。

根据式(20), 可对约束函数进行修正, 并得到新的能量函数为

$$\begin{aligned} E' &= \frac{a}{2} E'_1 + \frac{b}{2} E'_2 + c E'_3 \\ &= \frac{a}{2} \sum_{x'=1}^{N_w} \sum_{i'=1}^{D_w} \sum_{j'=1, j' \neq i'}^{D_w} V_{x'i'} V_{x'j'} + \\ &\quad \frac{a}{2} \left(\sum_{x'=1}^{N_w} \sum_{i'=1}^{D_w} V_{x'i'} - N_w \right)^2 - \\ &\quad \frac{b}{2} \sum_{x'=1}^{N_w} \sum_{i'=1}^{D_w} \sum_{y'=1}^{N_w} \sum_{j'=1}^{D_w} C_{x'i'y'j'} V_{x'i'} V_{y'j'} - \\ &\quad c \sum_{x'=1}^{N_w} \sum_{i'=1}^{D_w} S_{x'i'} V_{x'i'} \end{aligned} \quad (21)$$

将式(21)与式(5)进行比较, 得

$$W_{x'i'y'j'} = -a\delta_{x'y'}(1 - \delta_{i'j'}) + bC_{x'i'y'j'}(1 - \delta_{x'y'})(1 - \delta_{i'j'}) \quad (22)$$

$$I_{x'i'} = aN_w + cS_{x'i'} \quad (23)$$

3.4 算法的实现

(1) 初始化

设定 a, b, c 的初值;

由式(22)和式(23)分别计算 W_{xijy} 和 I_{xi} ;

设定神经元的初始状态 $u_{xi}(0)$ 和 $V_{xi}(0)$ 。由于神经元的输出 V_{xi} 反映了像点 x 和 i 的匹配几率, 在最初没有任何已知的先验信息的条件下, 左像点 x 与所有可能对应点 $i (x + d_{\min} \leq i \leq x + d_{\max})$ 的匹配几率是相同的, 因此有

$$V_{xi}(0) = 1/D_w \quad (24)$$

将式(24)代入式(4), 可得

$$u_{xi}(0) = (-u_0/2) \ln(D_w - 1) \quad (25)$$

当所有神经元具有相同的初值时, 整个网络将处于不稳定的平衡态, 而无法找出最小的能量

点^[10]。为了克服这一缺陷,在神经元初始值设定时往往加上一个较小的随机数^[5]:

$$u_{xi}(0) = (-u_0/2)\ln(D_w - 1) + \text{ran}(0.1 \times (-u_0/2)\ln(D_w - 1)) \quad (26)$$

其中,ran(x)为取值在 -x 和 x 间的随机数。

设定 $n=0$,该参数用于表示某一时刻神经元的输出值大于预定阈值 θ 的总个数;

(2) 在时刻 t ,随机选择神经元 n_{xi} ,利用式(3)

和式(4)计算 n_{xi} 在时刻 $t+1$ 的状态值,即 $u_{xi}(t+1)$ 和 $V_{xi}(t+1)$;

(3) 若 $V_{xi}(t+1) > \theta$,则 $n=n+1$;

(4) 若 n 未达到预定值 Z_0 ,则返回第 2 步;否则结束,并做如下处理

for $x=1:N_w$

for $i=1:D_w$

If $V_{xi}(m_{fin}) \geq \theta$ Then $disp(x)=i$

end

end

$disp(x)$ 表示像点 x 的视差, m_{fin} 表示稳态。

4 实验结果

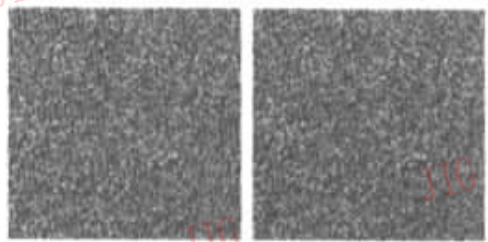
进行 3 组实验,一组实验对 3.4 节的方法进行检验,并进行了改进,取得了较好的效果。在改进方

法的基础上,进行另外两组实验,一组对常见的实际像对进行匹配,验证算法的有效性。另一组用于检验本算法的匹配精度。

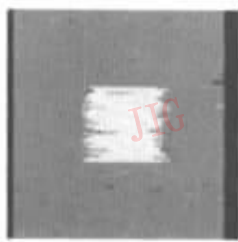
三组实验采用的网络参数为: $a=10, b=5, c=20, \theta=0.9, u_0=0.02$ 。 a, b, c, θ 取值直接影响了网络进入稳态的速度,进而决定了所得解的质量,一般来说,若 a, b, c 取值过大或 θ 取值较小时,神经元经过几次状态更新就能达到设定的要求,但解的质量不高,即错误匹配发生的几率较大;相反地,若 a, b, c 取值较小或 θ 取值过大时,可以获得较优的解,但网络需要很长时间才能实现稳定,甚至无法达到稳态;在参考文献[5]、[6]以及多次比较的基础上,确定了上述参数值。另外, $\lambda=0.3, G_0=0.05, Z_0$ 的取值与匹配的像点数目有关,一般取 $Z_0=N_w$ 。

4.1 实验 1

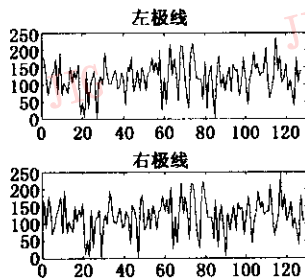
图 4(a)所示为 128×128 的随机点图对,其视差范围如表 1 所示,图 4(b)表示用 3.4 节算法得到的匹配结果。显然,这个结果不令人满意。这与 HNN 的本身特性有关。由于 HNN 为一个梯度系统,它只具有局部搜索的能力,因此解的质量与初始状态有很大的关系。图 4(c)和(d)对此进行了验证。图 4(c)所示为从图 4(a)中任意抽取的一对共轭极线,图 4(d)为利用上述算法从不同初始状态得到的 8 组视差曲线,它们之间的差异是明显的。



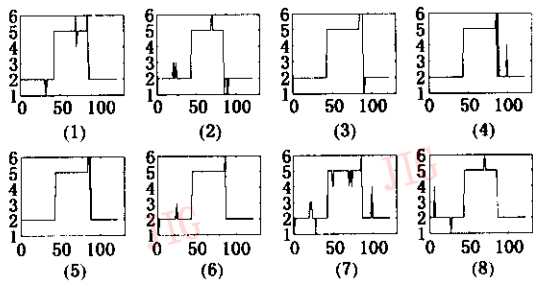
(a) 随机点图



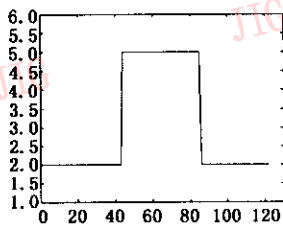
(b) 未改进的匹配结果



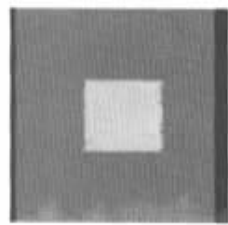
(c) 任意抽取的一对共轭极线



(d) 多次匹配的视差曲线



(e) 改进后的视差曲线



(f) 改进后的匹配结果

图 4 随机点图的匹配结果

为了消除初始状态的选取对最后结果的影响,采用一个简单的处理办法:从不同的初始状态进行多次匹配操作,然后从得到的一组次优解中找出最优解。具体操作是:若在 M 次匹配中,左像点 x 得到的对应解为 $i_l (l=1, 2, \dots, D_w)$ 的次数记为 $N(x, i_l)$, 且有

$$\sum_{l=1}^{D_w} N(x, i_l) = M \quad (27)$$

则像点 x 的最终对应点 $i_{l_{opt}}$ 满足

$$N(x, i_{l_{opt}}) = \max\{N(x, i_1), N(x, i_2), \dots, N(x, i_{D_w})\} \quad (28)$$

像点 x 的视差记为 $dist(x) = i_{l_{opt}}$ 。

图 4(e) 为采用改进方法对图 4(c) 进行匹配后得到的视差曲线图,图 4(f) 为对整个像对进行匹配的视差图,改进后的结果明显优于图 4(b)。

4.2 实验 2

图 5 所示为几幅常见的立体像对,其中(a)为合成图,(b)~(d)为实际景物图,各个像对的尺寸以及视差范围如表 1 所示。(e)~(h)为利用本文算法得到的视差图(各个像对进行匹配所用时间见表 1,

表 1 像对进行匹配运算的参数值

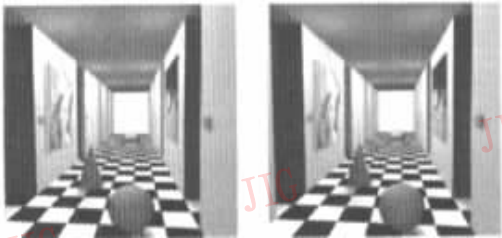
	图像大小	视差范围(pixels)	匹配时间(s)
随机点图	128×128	[1 6]	5.2
corridor	256×256	[0 10]	39.7
pentagon	512×512	[-13 12]	558.1
shrub	256×240	[1 9]	36.5
parking meter	512×480	[0 23]	513.2

计算机主频为 P4-2.4G。由匹配结果,得出如下几点结论:

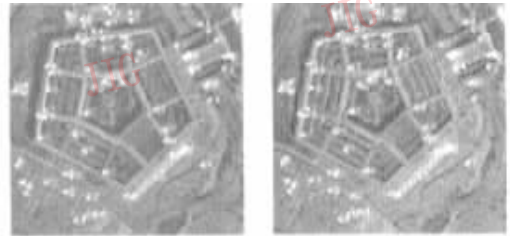
(1) 算法能够将景物中各主要物体的深度差别较好地地区分开,如 shrub 图中的路牌、灌木丛、房屋(图 5(g))以及 parking meter 图中的花卉和汽车(图 5(h))等,又如 pentagon 图中的五角大楼与周边街区(图 5(f))等。

(2) 物体的外形轮廓可以在得到的视差图中较好地恢复,如 corridor 图中的小球和圆锥物(图 5(e))和 pentagon 图中的五角大楼(图 5(f))。

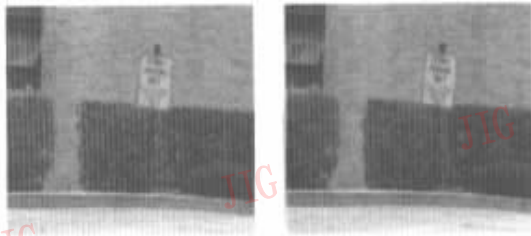
(3) 景物中一些细微物体的深度差也能在视差图中得到较好的反映,如 corridor 图中走廊尽头的方形物,以及 shrub 图中的花卉等。



(a) corridor 原始像对



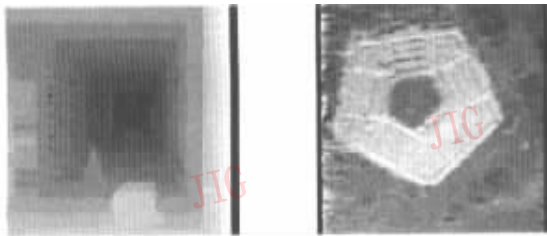
(b) pentagon 原始像对



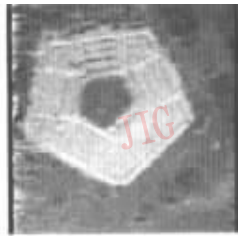
(c) shrub 原始像对



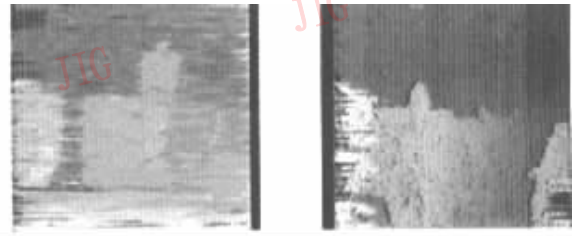
(d) parking meter 原始像对



(e)



(f)



(g)



(h)

图 5 匹配结果

存在的主要问题是视差图不够平滑,且存在相当数量的水平条纹(如图 5(g))。产生这种情况的主要原因是:建立平滑约束函数时,只考虑了对应极线上邻点之间的相容性要求,而对于相邻极线的约束信息,则未予考虑,因此匹配得到的最终结果并非整体最优,相邻极线间出现纵向视差不连续现象,也就在所难免。为了解决上述问题,一个可行的办法是在能量函数中引入极线间的平滑约束,使得到视差结果在水平和垂直方向均得到平滑,从而实现全局最优。

4.3 实验 3

这组实验用于验证算法的匹配精度,描述错误匹配率的函数为

$$T_M(i) = \frac{N_D(i)}{N_A}, i = 0, 1, 2, \dots \quad (29)$$

其中, N_A 表示用于匹配的图像点数, $N_D(i)$ 表示算法得到的视差值与理想视差值相差为 i 个像素的点



(a) corridor 的立体视差图



(b) 利用 Canny 算子对 pentagon 像对进行边缘检测的结果

图 6 精度测试中所用到的图像

图 6(b)为利用 Canny 算法对 pentagon 像对进行边缘检测的结果。利用基于特征的 HNN 方法^[5]对该图中的边缘点进行匹配,可以得到一组对应关系,由于特征点数过多(左、右图分别有 458 812 和 460 226 个特征点),受篇幅所限,无法将所有点的视差值一一列出。可以指出的是,(1)上述匹配采用的参数值均可以从文献^[5]中查到;(2)为了加快匹配速度,将图 6(b)中图像区域划分为 16×16 个方块,每块的大小为 32×32 ,匹配操作在每块独立进行。

将本算法以及上述方法得出的结果作匹配误差比较,其中,本文方法所得的视差结果如图 5(f)所示。为了使比较具有一定的客观性以及准确性,预先从 2 幅图中找出 78 个对应关系已知的点(点的位置在图中用白圈标出,如图 5(b)所示),记为 $(p_l(i), p_r(i)), i = 1, 2, \dots, 78, p_l(i), p_r(i)$ 表示左右对应像点坐标。利用已知的匹配结果,可以直接得出 2 种方

数。显然, $T_M(0)$ 表示该次匹配运算的正确率。

图 6(a)所示为 corridor 合成图对的理想深度图,图中各点的视差值可由理论计算得到,将其与图 5(e)对应的视差值进行比较后,利用式(29)求得匹配的误差率,如表 2 所示。

表 2 合成像对经过匹配后的误差分布

$T_M(i)$	$i=0$	$i=1$	$i=2$	$i=3$	$i \geq 3$
随机点图	0.992	0	0.008	0	0
corridor 图	0.784	0.105	0.022	0.068	0.021

从结果可知,78.4%的点实现了正确匹配,匹配误差在 3 个像素以上的仅有 2.1%的图像点,而其余点的匹配误差均在 3 个像素以内,这表明,利用本文算法能够基本恢复出场景中的深度信息。采用相同的方法,对图 4(a)的随机点图的所得的匹配结果也进行了误差测试,结果表明(见表 2),99.2%的像点得到了正确匹配。

法在 $I_1 = \{p_l(i), i = 1, 2, \dots, 78\}$ 的视差值,将其与理想值进行比较后,求得匹配误差的分布结果,如表 3 所示。必须指出的是,基于特征方法得到的在 I_1 上的视差并非全是整数,为了便于比较,对其进行了取整操作。

表 3 78 对像点在 2 种匹配运算后的误差分布

$T_M(i)$	$i=0$	$i=1$	$i=2$	$i=3$	$i \geq 3$
本文算法	0.705	0.077	0.103	0.064	0.051
特征匹配方法	0.667	0.115	0.064	0.026	0.128

由表 3 可知,利用本算法共有 70.5%的点实现了正确匹配,而基于特征方法的正确匹配率为 66.7%,两者相差不大。另外,本算法整体匹配误差较小,匹配误差多于 3 个像素的点仅有 4 个,而后一种方法却有 10 个之多。分析结果表明,匹配错误较大的点基本分布在图像的左下部分,该区域相似的图像细节多重出现,因而难以找出准确的对应关系。

5 结 论

新的基于 Hopfield 网络的立体匹配算法,与传统方法相比,具有一定的新意。

(1) 本算法采用普通的像素点作为匹配单元,好处在于能够获得稠密的深度图,且避免了传统的特征提取问题。然而,一般的观点认为,由于立体投影变换过程中并不保持光度不变性,并且考虑诸如光照变化、噪声等干扰因素的影响,基于像素点的立体视觉方法“必然是行不通的”^[11]。对此,也有人表示了不同的看法^[9,12],特别是 Gennert 在进行深入研究后明确指出了基于灰度匹配方法的可行性^[13],而本文的实验结果有力地支持了这一观点。

(2) 利用优化方法求解对应问题,是目前立体匹配研究的一个趋势。可以说,现在最好的匹配方法,均采用了优化手段^[2,14]。利用 HNN 来解决对应问题,早已有之,但传统的方法不能直接得到稠密的深度图,因而无太大的实用价值。本文提出的方法,初步解决了这个难题,是一个有益的尝试。然而,算法所得的匹配结果并非全局最优的,这与采取的匹配策略有关,也与采用的网络拓扑结构有关,由于 2D HNN 为一平面结构,因而无法刻划整个 2D 视差图景演化过程。为了得到一个满足整体最优的匹配结果,关键在于构造一个 3D Hopfield 网络,并使该网络经过动态更新后得到的稳态对应于最优的视差曲面,这也是下一步工作的重点。

参 考 文 献

- 1 Kim Y, Aggarwal J. Positioning three-dimensional objects using stereo images [J]. *IEEE Journal of Robotics and Automation*, 1987, **3**(4): 361~373.
- 2 Sun Changming. Fast stereo matching using rectangular subregioning and 3D maximum surface techniques [J]. *International Journal of Computer Vision*. 2002, **47** (1/2/3): 99~117.
- 3 Hopfield J, Tank. D W. Computing with neural circuits: A model [J]. *Science*, 1986, **233**(4764): 625~633.
- 4 Nasrabadi N M, Li W, Epranian B G, *et al.* Use of Hopfield network for stereo vision correspondence [A]. In: *IEEE International Conference on Systems, Man and Cybernetics* [C]. Cambridge MA, USA, 1989, **2**: 429~432.

- 5 Nichani S. Solving the correspondence problem using a Hopfield network [A]. In: *IEEE International Conference on Neural Networks* [C]. Orlando, Florida, USA, 1994, **6**: 4107~4112.
- 6 Huang Chia-Horng, Wang Jung-Hua. Stereo correspondence using Hopfield network with multiple constraints [J]. *IEEE International Conference on Systems, Man, and Cybernetics*, 2000, **2**(2): 1518~1523.
- 7 Burt P, Julesz. A Disparity Gradient Limit for Binocular Fusion [J]. *Science*, 1980, **208**: 615~617.
- 8 Pollard S B, Mayhew J E W, Frisby J P. PMF: A stereo correspondence algorithm using a disparity gradient limit [J]. *Perception*, 1985, **14**(4): 449~470.
- 9 Cox I J, Hingorani S L, Rao S B. A maximum likelihood stereo algorithm [J]. *Computer Vision and Image Understanding*, 1996, **63**(3): 542~567.
- 10 Lin W, Delgado-Frias J, Pechanek G G, *et al.* Impact of energy function on a NN model for optimization problems [A]. *International Joint Conference on Neural Networks'94* [C], Lake Buena Vista (Orlando), Florida, USA, 1994: 4518~4523.
- 11 Frisby J, Pollard S. Computational issues in solving the stereo correspondence problem [A]. In: Landg M S, Movshon J A editors, *Computational Models of Visual Processing* [M]. Cambridge, MA, USA: MIT Press, 1991: 331~357.
- 12 Belhumeur P. A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion [A]. In: *Proc. Int. Conf. on Computer Vision* [C]. Berlin, 1993.
- 13 Gennert M A. A Computational Framework for Understanding Problems in Stereo Vision [M]. Cambridge Massachusetts: MIT Press, 1987.
- 14 Roy S. Stereo without epipolar lines: A maximum-flow formulation [J]. *International Journal of Computer Vision*, 1999, **34**(2): 147~161.



胡海峰 1977 年生,现为中山大学电子与通信工程系博士生。主要研究方向为图像处理与机器视觉、神经网络、模式识别。

熊银根 1962 年生,副教授。主要研究领域为人工智能与模式识别,计算机视觉。